



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/607,370	06/30/2000	Reiner Kraft	ARC9-2000-0046-US1	8705

23334 7590 02/24/2005

FLEIT, KAIN, GIBBONS, GUTMAN, BONGINI
& BIANCO P.L.
ONE BOCA COMMERCE CENTER
551 NORTHWEST 77TH STREET, SUITE 111
BOCA RATON, FL 33487

EXAMINER

BURGE, LONDRA C

ART UNIT

PAPER NUMBER

2178

DATE MAILED: 02/24/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/607,370

Applicant(s)

KRAFT ET AL.

Examiner

Londra C Burge

Art Unit

2178

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 01 February 2005.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-20 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-20 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
 - ☐ Certified copies of the priority documents have been received in Application No. _____.
 - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____.
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: _____.

DETAILED ACTION

1. This after is responsive to communications: After Final Amendment filed 2/1/2005.
2. Claims 1-20 are pending and claims 1, 14 and 20 are independent claims.
3. The Final Rejection using Raman, King et al., Meyerzon et al. and Meyerzon et al. have been withdrawn based on arguments received from the applicant.
4. This office action is made Non-Final and new grounds of rejections have been presented.

Claim Rejections - 35 USC § 103

5. **The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:**

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

6. **Claims 1-3, 14-16, 20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (herein after Meyerzon) U.S. Patent No. 6,638,314 B1 filed 6/26/1998 in view of Lawrence et al. (herein after Lawrence) U.S. Patent No. 6,289,342 B1 filed 5/20/1998.**

In regard to independent claim 1, Meyerzon discloses *retrieving a web document at an address and extracting contents of the web document for rendering an intermediate dynamically constructed in-memory web page representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser* (Meyerzon Col 7 Lines 60-65 and Col 8 Lines 15-20 i.e. web crawler program searches remote server computers connected to the network for electronic documents and retrieves electronic documents and

Art Unit: 2178

associated data and a browser displays documents to a user); *loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory web page representation* (Meyerzon Col 8 Lines 26-35 i.e. the client computer transmits data to a search engine, the search engine examines its associated index to find documents and returns the documents which are secondary documents and lists the documents for the user to view), *wherein the secondary documents include one or more images with textual content embedded therein* (Meyerzon Col 9 Lines 44-50 i.e. visual element m include text and hyperlink to an image); *analyzing and summarizing the in-memory web page representation to produce a text map for the web page document of the textual contents* (Meyerzon Col 10 Lines 13-16 i.e. passes the lists of properties and text to the indexing engine and the indexing engine creates an index, which is used by the search engine in subsequent searches).

Meyerzon does not specifically mention *using optical character recognition* on the images to extract textual content for adding to the textual map for the web page document. However, Lawrence mentions extracting data using optical character recognition (Lawrence Col 7 Lines 51-56 i.e. conversion to electronic form by use of OCR). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Lawrence to Meyerzon, providing Meyerzon the benefit of extracting content from a document using OCR, which is quicker the typing out an entire document manually by hand.

In regard to dependent claim 2, which depends on claim 1, Meyerzon discloses *wherein the retrieving the web document at an address further comprises retrieving a document at an address selected from the group of addresses consisting of a nodal address, a network address, a URL and equivalents* (Meyerzon Col 21 Lines 1-11 i.e. a request to retrieve a list of

Art Unit: 2178

electronic documents and retrieving a set of document address specifications corresponding the electronic documents).

In regard to dependent claim 3, which depends on claim 1, Meyerzon discloses *wherein the one or more images with textual content embedded therein include at least one of an in-line GIF image and an in-line JPEG image.* (Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page and it is well known in the art the images displayed on web pages can be a gif and jpeg image).

In regard to dependent claim 7, which depends on claim 1, Meyerzon discloses *initializing a first list with seed values* (Meyerzon Col 17 Lines 25-26 i.e. assigning a current crawl number to the current web crawl); *checking if there are any URLs to be processed and in response that any URL exists to be processed then performing the follow sub-steps of* (Meyerzon Col 17 Lines 28-29 i.e. determine whether an electronic document has been retrieved): *determining if a URL is in a second list; and in response that a URL is not in the second list then performing the following sub-steps of: inserting the URL into the first list; scheduling the URL for crawling; crawling the URL when scheduled to do so; removing the URL from the first list after the scheduled crawling; entering the URL into the second list* (Meyerzon Col 9 Lines 64 and Col 10 Lines 1-11 i.e. history map checks each hyperlink URL to determine if it is already listed in the history map, if not the URLs are added and are marked as not being crawled and added to the transaction log. The history map includes a number crawled and number modified data); *and repeating the checking step until there are no more URLs to be processed; where if the determining step determines that the URL is in the second list then repeating the checking*

Art Unit: 2178

step until there are no more URLs to be processed. (Meyerzon Col 12 Lines 1-17 i.e. retrieves and processed a URL until there are none left in the transaction log)

In regard to dependent claim 8, which depends on claim 7, Meyerzon discloses *wherein the sub-step of initializing a first list with seed values further includes the list being a URL pool. (Meyerzon Col 7 Lines 65-67 i.e. retrieving a processing URLs from the transaction log)*

In regard to dependent claim 9, which depends on claim 7, Meyerzon discloses *wherein the sub-step of determining if a URL is in a second list further includes the second list being a visited pool. (Meyerzon Figure 4 shows a column indicating the number crawled and modified)*

In regard to dependent claim 10, which depends on claim 7, Meyerzon discloses *wherein the sub-step of crawling further comprises the sub-steps of: issuing an HTTP command to a web server named in the URL; receiving contents of an HTML page as a result of the issued HTTP command; and passing on the contents of the HTML page to a Page Rendering subroutine. (Meyerzon Col 8 Lines 26-35 i.e. the client computer transmits data to a search engine, the search engine examines its associated index to find documents and returns the documents which are secondary documents and lists the documents for the user to view)*

In regard to dependent claim 11, which depends on claim 10, Meyerzon discloses *receiving the contents of the HTML page in the Page Rendering subroutine; building an in-memory representation of a layout for the HTML page and if more data is needed to properly form the representation, then performing the sub-steps of (Meyerzon Col 7 Lines 60-65 and Col 8 Lines 15-20 i.e. web crawler program searches remote server computers connected to the*

Art Unit: 2178

network for electronic documents and retrieves electronic documents and associated data and a browser displays documents to a user); *requesting additional web-based information; gathering this additional web-based information; inserting any URLs associated with this additional web-based information into the second list and a URL cache* (Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page); *building a final amended representation; and forwarding the final amended representation to an Extraction subroutine; wherein, if no more data is needed to properly form the in-memory representation, then forwarding the in-memory representation to the Extraction subroutine.* (Meyerzon Col 16 Lines 32-44)

In regard to dependent claim 12, which depends on claim 11, Meyerzon discloses *accessing a set of memory structures of the Page Renderer* (Meyerzon Col 6 Lines 23-60 i.e. accessing local and remote memory devices); *copying a text portion of the structures into a text map* (Meyerzon Col 15 Lines 15-16 i.e. copying all of the history map entries into the transaction log as entries); *inspecting any in-line GIF and JPEG image references in the memory structures* (Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page and it is well known in the art the images displayed on web pages can be a gif and jpeg image); *extracting alternate text attributes* (Meyerzon Col 5 Lines 7-8 i.e. extracting data from each of the retrieved documents); *adding the alternate text attributes to a text map* (Meyerzon Col 2 Lines 48-51 i.e. information from the electronic document retrieved from the web crawl is stored in an index); *extracting text content from the GIF and JPEG images; adding text content from the images to the text map* (Meyerzon Col 9 Lines 37-46 i.e. an image is retrieved to display on a web page and it is well known in the art the images displayed on web pages can be a gif and jpeg image Col 5 Lines 7-8 i.e. extracting data from each of the retrieved documents Col 2 Lines 48-51 i.e.

Art Unit: 2178

information from the electronic document retrieved from the web crawl is stored in an index); *and forwarding the text map to a Page Summarizer subroutine.* (Meyerzon Col 9 Lines 64 and Col 10 Lines 1-11 i.e. history map checks each hyperlink URL to determine if it is already listed in the history map, if not the URLs are added and are marked as not being crawled and added to the transaction log. The history map includes a number crawled and number modified data)

Meyerzon does not specifically mention *invoking an optical character recognition engine*; analyzing any in-line GIF and JPEG images *using the optical character recognition engine* for text content. However, Lawrence mentions extracting data using optical character recognition (Lawrence Col 7 Lines 51-56 i.e. conversion to electronic form by use of OCR). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Lawrence to Meyerzon, providing Meyerzon the benefit of extracting content from a document using OCR, which is quicker the typing out an entire document manually by hand.

In regard to dependent claim 13, which depends on claim 12, Meyerzon discloses *receiving a text map from the Page Extractor subroutine; processing the text map in an application-specific manner* (Meyerzon Col 2 Lines 48-51 i.e. information from the electronic document retrieved from the web crawl is stored in an index to begin the routine); *applying data extraction patterns to the text map* (Meyerzon Col 5 Lines 7-8 i.e. extracting data from each of the retrieved documents); *translating resultant data from the applying step; forwarding any URLs present in the text map to a manager subroutine; and forwarding any extracted data and metadata to application logic.* (Meyerzon Col 9 Lines 64 and Col 10 Lines 1-11 i.e. history map checks each hyperlink URL to determine if it is already listed in the history map, if not the URLs

Art Unit: 2178

are added and are marked as not being crawled and added to the transaction log. The history map includes a number crawled and number modified data)

In regard to independent claims 14 and 20, claims 14 and 20 in addition to the following reflect similar subject matter claimed in claim 1 and are rejected along the same rationale. (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction and Col 20 Lines 23-24 i.e. a system for retrieving stored information)

In regard to dependent claim 15, which depends on claim 14, claim 15 in addition to the following reflect similar subject matter claimed in claim 2 and are rejected along the same rationale. (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction)

In regard to dependent claim 16, which depends on claim 14, claim 16 in addition to the following reflect similar subject matter claimed in claim 3 and are rejected along the same rationale. (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction)

7. **Claims 4-6 and 17-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (herein after Meyerzon) in view of Lawrence et al. (herein after Lawrence) as applied to claim 1 and in further view of Hobbs U.S. Patent No. 6,523,022 B1 filed 7/7/1999.**

In regard to dependent claim 4, which depends on claim 1, Meyerzon does not specifically mention *wherein the loading secondary documents further comprises the loading of secondary documents including one or more Java applets with textual content embedded therein.*

Art Unit: 2178

However, Hobbs mentions that Java applets are used (Hobbs Col 28 Line 35). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Hobbs to Meyerzon, providing Meyerzon the benefit of using Java Applets for web pages in the process of searching the web documents because Java Applets are compatible with many web pages and browsers.

In regard to dependent claim 5, which depends on claim 1, Meyerzon does not specifically mention *wherein the loading secondary documents further comprises the loading of secondary documents including web documents selected from the group of documents consisting of in-line frames, frames, and equivalents*. However, Hobbs mentions that frames and in-line frames are used (Hobbs Col 7 Lines 63 through Col 8 Lines 1-34). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Hobbs to Meyerzon, providing Meyerzon the benefit of using frames and in-line frames for easy viewing for the user.

In regard to dependent claim 6, which depends on claim 4, Meyerzon does not specifically mention *wherein the loading secondary documents further comprises the loading of secondary documents including one or more Java Script components with textual content embedded therein*. However, Hobbs mentions that Java applets are used (Hobbs Col 28 Line 35). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Hobbs to Meyerzon, providing Meyerzon the benefit of using Java Scripts for web pages in the process of searching the web documents because Java Scripts are compatible with many web pages and browsers.

In regard to dependent claim 17, which depends on claim 14, claim 17 in addition to the following reflect similar subject matter claimed in claim 4 and are rejected along the same

Art Unit: 2178

rationale. (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction)

In regard to dependent claim 18, which depends on claim 14, claim 18 in addition to the following reflect similar subject matter claimed in claim 5 and are rejected along the same rationale. (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction)

In regard to dependent claim 19, which depends on claim 17, claim 19 in addition to the following reflect similar subject matter claimed in claim 6 and are rejected along the same rationale. (Meyerzon Col 20 Lines 13-14 i.e. computer readable medium having computer executable instruction)

Response to Arguments

8. Applicant's arguments, filed 2/1/2005, with respect to the rejection(s) of claim(s) 1-20 under 35 USC 103(a) have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of Meyerzon et al., Lawrence et al., and Hobbs.

Regarding claims 1, 14 and 20. The applicant argues that the original rejection of Raman in view of King does not teach rendering a web document and extracting content from the document, loading a secondary document, analyzing and summarizing the web page representation to produce a text map and using OCR to extract data for the web page document (Pages 14-17). The examiner agrees, however in the new grounds of rejections, Meyerzon discloses *retrieving a web document at an address and extracting contents of the web document*

Art Unit: 2178

for rendering an intermediate dynamically constructed in-memory web page representation of the web document at a hub processing unit which is formatted as if displayed for viewing on an end-user's web browser (Meyerzon Col 7 Lines 60-65 and Col 8 Lines 15-20 i.e. web crawler program searches remote server computers connected to the network for electronic documents and retrieves electronic documents and associated data and a browser displays documents to a user); *loading secondary documents associated with the web document in order to render the secondary documents as part of the in-memory web page representation* (Meyerzon Col 8 Lines 26-35 i.e. the client computer transmits data to a search engine, the search engine examines its associated index to find documents and returns the documents which are secondary documents and lists the documents for the user to view), *wherein the secondary documents include one or more images with textual content embedded therein* (Meyerzon Col 9 Lines 44-50 i.e. visual element m include text and hyperlink to an image); *analyzing and summarizing the in-memory web page representation to produce a text map for the web page document of the textual contents* (Meyerzon Col 10 Lines 13-16 i.e. passes the lists of properties and text to the indexing engine and the indexing engine creates an index, which is used by the search engine in subsequent searches).

Meyerzon does not specifically mention *using optical character recognition* on the images to extract textual content for adding to the textual map for the web page document. However, Lawrence mentions extracting data using optical character recognition (Lawrence Col 7 Lines 51-56 i.e. conversion to electronic form by use of OCR). It would have been obvious to one of ordinary skill in the art at the time of the invention, to apply Lawrence to Meyerzon,

Art Unit: 2178

providing Meyerzon the benefit of extracting content from a document using OCR, which is quicker the typing out an entire document manually by hand.

Regarding claim 4, the applicant argues that the original rejection of Raman in view of King does not teach of Java applets used (Page 17). The examiner agrees, however in the new grounds of rejections, Hobbs mentions that Java applets are used (Hobbs Col 28 Line 35).

Conclusion


Any inquiry concerning this communication or earlier communications from the examiner should be directed to Londra C Burge whose telephone number is (571) 272-4122.

The examiner can normally be reached on 8:30am to 5:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Stephen Hong can be reached on (571) 272-4124. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

LCB
2/18/2005


CESAR B PAULA
PRIMARY EXAMINER
AU 2178